FUNDAÇÃO GETULIO VARGAS

EPGE

Escola de Pós-Graduação em Economia

Ensaios Econômicos

Escola de
Pós Graduação
em Economia
da Fundação
Getulio Vargas

Nº 515

ISSN 0104-8910

O método generalizado dos momentos(MGM): Conceitos básicos

Renato G. Flôres Jr

Novembro de 2003

O MÉTODO GENERALIZADO DOS MOMENTOS (MGM): CONCEITOS BÁSICOS

Renato G. Flôres Jr.

(Setembro, 2003)

RESUMO

O presente texto desenvolve, com fins didáticos, as propriedades básicas do Método Generalizado dos Momentos (MGM). Faz parte de obra (livro) em elaboração.

ABSTRACT

This text presents the basic properties of the Generalised Method of Moments (GMM). It makes for a chapter of a work (textbook) in progress.

Algumas abreviações e convenções

Abreviações

iid ou i.i.d. - *(diz-se de uma amostra com observações) independentes e identicamente distribuídas*

LfGN - *(uma) lei fraca dos grandes números*

MQG - *(o método ou o estimador de) mínimos quadrados generalizados*

MQO - *(o método ou o estimador de) mínimos quadrados ordinários*

MQ2 - *(o método ou o estimador de) mínimos quadrados em dois estágios*

TCL - *(um) teorema central do limite*

varass - *variância assintótica*

VI - *(o método ou o estimador de) variáveis instrumentais*

Convenções

jacobianos - dada uma função $g:R^p \rightarrow R^q$, denota-se por g(x) o vetor *coluna* do R^q correspondente ao ponto $x \in R^p$. O jacobiano $[\partial g/\partial x]$ será sempre, por convenção, uma matriz q x p.

vetorizações – o operador vec, aplicado a uma matriz nxm A, vecA, a transforma, salvo menção explícita em contrário, em um vetor coluna de dimensão nm, obtido pelo empilhamento sucessivo das *colunas* de A.

Capítulo 1

O MÉTODO GENERALIZADO DOS MOMENTOS (MGM): CONCEITOS BÁSICOS

1. Introdução: uma ironia histórica.

O método generalizado dos momentos (MGM) tem a sua origem - como muitas outras idéias em estatística - no trabalho precursor de Lord Karl Peason (↑1857, Londres; ↓1936, Londres). Pearson havia construído o seu sistema de famílias de distribuições de probabilidade e necessitava um procedimento automático para, uma vez escolhida a família apropriada, estimar os parâmetros da particular curva (ou função de densidade) que melhor se ajustaria aos dados. Como tais parâmetros eram, geralmente, relacionados aos momentos da distribuição através de funções algébricas não muito complicadas, ele propôs que se igualasse os valores dos momentos amostrais aos momentos teóricos, conforme descritos por tais funções. Desse modo, indo-se até o momento de ordem igual ao número de parâmetros a estimar, se obtinha um sistema de tantas equações quanto incógnitas - os parâmetros - a resolver.

Exemplificando, supondo-se que a distribuição fosse individualizada pelos três parâmetros α, β e γ e que existissem três funções g_i , $i=1,2,3$, unindo-os aos momentos até à ordem 3, isto é,

$$g_i(\alpha,\beta,\gamma) = \mu_i = EX^i \quad , \quad i=1,2,3 \quad ;$$

dada uma amostra i.i.d. $X_1, X_2, ..., X_t, ..., X_T$, de tamanho T, a sua proposta era usar como estimadores a solução do sistema

$$g_i(\alpha,\beta,\gamma) = (1/T) \sum_t X_t^i \quad , \quad i=1,2,3 . \tag{1.1}$$

A discussão entre as vantagens e a adequação desta técnica, relativamente às idéias que o (então) jovem Ronald A. Fisher (↑1890, Londres; ↓1962, Adelaide, Austrália) propunha – particularmente o método da máxima verossimilhança (MV), que ele desenvolveu e adotaria na estimação de parâmetros -, custou a este o seu emprego na Universidade de Londres, indo finalmente parar, em 1919, na estação experimental agrícola de Rothamsted, onde se tornou célebre. Devido à influencia que ganharam as idéias de Fisher, o método dos momentos - e todo o sistema de famílias de distribuições a ele ligado - permaneceu em razoável ostracismo, enquanto a máxima verossimilhança ganhou um respeitável *pedigree* entre os “estatísticos matemáticos”.

É irônico notar que, após mais de trinta anos de sucesso da MV, o método dos momentos é “redescoberto” e trazido novamente ao panteão - sob uma roupagem bem mais geral é verdade – graças ao artigo fundamental de Hansen (1982) e aos trabalhos pioneiros de Burguete (1980), Gallant (1977), Gallant and Jorgenson (1979) e Manski (1988), entre outros. Ressalte-se ainda que, as propostas de Pearson sobre como “identificar” inicialmente a família adequada, utilizando-se fortemente o histograma, podem também ser consideradas como uma postura (ainda que rudimentar em termos modernos) de abordagem não-paramétrica. Ao menos, refletem uma atitude mais flexível e “respeitosa” aos dados, do que a imposição muitas vezes arbitrária de uma verossimilhança específica.

2. Um arcabouço geral – os estimadores de extremos.

O MGM pode ser estudado como um caso particular de uma classe mais geral de estimadores - os chamados estimadores de extremos ou estimadores M[1] -, no entanto, como de hábito quando se almeja excessiva generalidade, esta abordagem é boa para a obtenção de certas propriedades e demasiado ampla para realçar outras questões, sobretudo em um primeiro contato com o método. Explorando, portanto, o aspecto positivo, nesta seção apresentamos alguns conceitos básicos relativos aos estimadores M, que ajudam a ilustrar aspectos importantes dos resultados apresentados na próxima seção.

Seja o vetor aleatório W, r-dimensional, e um parâmetro $\theta \in \Theta$, sendo Θ um subconjunto - por ora sem nenhuma estrutura particular - do R^p . Conhecida uma amostra $\{W_t\}_{1 \leq t \leq T}$, o estimador de extremo de θ é definido como a solução de um programa de maximização (ou minimização) de uma função $Q_T(\{W_t\}_{1 \leq t \leq T} ; \theta)$, suposta conhecida. Esta abordagem engloba diversos métodos conhecidos, como a MV, os mínimos quadrados não-lineares, o MGM e os estimadores de distância mínima. Sob ela, é possível obter resultados a respeito da consistência e da normalidade assintótica dos estimadores. Naturalmente, ambos pressupõem resolvidas as questões de existência e identificação. Para tal, é comum exigir continuidade da Q_T com relação a θ e que Θ seja compacto. A hipótese de compacidade é um pouco desagradável, pois, como sabido, implica, nos espaços euclideanos, em que Θ seja limitado. Uma saída prática é assumir, nos casos concretos, que existam limites acima dos quais é totalmente improvável se situarem os valores de interesse. A identificação pode ser discutida junto com a consistência, como faremos a seguir.

---

[1] Deve-se ressaltar que não há unanimidade sobre a interpretação deste termo, alguns considerando que o M se refere a máximo ou mínimo e, portanto, sendo sinônimo de estimador de extremos, e outros dizendo

Obter consistência no contexto dos estimadores M é interessante porque a resposta é procurada na situação que interessa - a existência de um único extremo global - e não por vias alternativas, como as requerendo a existência de condições de primeira ordem, que, em possuindo solução, podem admitir raízes múltiplas, mesmo no caso de um único extremo. Para responder à questão sobre a consistência, investiga-se primeiro a convergência da função objetivo, obtendo-se o limite dos pontos extremos como o extremo do limite da função objetivo. Para que esta passagem seja válida, é comum impor-se que as convergências necessárias se dêem uniformemente no espaço do parâmetro[2]. Este tipo de hipótese é tão usado hoje em dia que justifica uma definição geral:

Definição 1. Seja $\{X_t(\theta) \, ; \, t=1,2,... \, ; \, \theta \in \Theta \subseteq R^p\}$ uma coleção de seqüências de vetores aleatórios de dimensão r, indexadas por θ, e $X^*(\theta)$ uma função que a cada $\theta \in \Theta \subseteq R^p$ associa um vetor aleatório r-dimensional. Diz-se que a coleção converge em probabilidade (quase certamente, em média quadrática, em distribuição) a $X^*(\theta)$, uniformemente em Θ,

se e somente se

a seqüência $\{\sup_\theta \left| X_t(\theta) - X^*(\theta) \right| \}$ converge em probabilidade (quase certamente, em média quadrática, em distribuição) a zero.

Note-se que a seqüência $\{\sup_\theta \left| X_t(\theta) - X^*(\theta) \right| \}$ é estocástica, pois o supremo, em cada instante t, é calculado para cada ponto do espaço amostral. A definição é sutil e

---

que o M se refere a "como o de Máxima verossimilhança", sendo o estimador, neste caso, uma versão particular do de extremos.

[2] Esta condição de uniformidade facilita as coisas, devido ao fato de que a escolha dos pontos é, de certa forma, arbitrária. A condição não é exatamente fundamental, podendo ser substituída por outra mais fraca - em termos de exigências sobre o comportamento dos limites em probabilidade para os pontos extremos - porém de redação bem mais inconveniente.

bastante forte, pois, no mesmo modo de convergência original, a seqüência dos supremos vai a zero, garantindo uma convergência uniforme em Θ *para cada ponto amostral*. Com este conceito, pode-se enunciar a hipótese que dá a condição de identificação:

Hipótese I (HI). A coleção $\{ Q_T(\{Wt\}_{1\leq t\leq T} ; \theta) \}$ converge em probabilidade uniformemente em Θ, a uma função determinística $Q: \Theta \to R^r$, contínua, tal que existe um valor $\theta_0 \in \Theta$, para o qual se $\theta \neq \theta_0$ então $Q(\theta) < Q(\theta_0)$.

A consistência agora pode ser obtida:

Proposição 1: Sob a hipótese HI, e se Θ é compacto, o estimador M é consistente.

Prova: É necessário provar que dada uma vizinhança aberta V, qualquer, de θ_0, tem-se

$$P[\theta_T \notin V] \to 0 \quad , \tag{2.1}$$

sendo θ_T o estimador M (ou o "ponto de ótimo" de Q_T) para uma amostra de tamanho T. Dado um $\varepsilon>0$, a Hipótese I implica que a probabilidade dos dois eventos seguintes tende a 1:

$$Q(\theta_T) > Q_T(\theta_T) - \varepsilon/3$$

$$Q_T(\theta_0) > Q(\theta_0) - \varepsilon/3 \quad .$$

Como, trivialmente,

$$Q_T(\theta_T) > Q_T(\theta_0) - \varepsilon/3 \quad ,$$

somando-se as três desigualdades tem-se que

$$P[Q(\theta_T) > Q(\theta_0) - \varepsilon] \to 1$$

ou, o que é o mesmo,

$$P\,[\,Q(\theta_0)\ -\ Q(\theta_T)\ >\varepsilon\,]\ \rightarrow\ 0 \qquad . \qquad (2.2)$$

Pode então parecer que este fato, junto à condição de identificação e à da continuidade da Q contidas na Hipótese I, seria suficiente para garantir a consistência do estimador. No entanto, necessita-se ainda o fato de que Θ seja compacto[3]. Dada uma qualquer V, vizinhança aberta de θ_0 , como $A = \Theta \cap V^c$ é compacto, a desigualdade abaixo faz sentido:

$$Q(\theta_0)\ -\ \sup_A Q(\theta)\ > 0 \qquad .$$

Tomando $\varepsilon = Q(\theta_0)\ -\ \sup_A Q(\theta)$, tem-se, por (2.2),

$$P\,[\,\sup_A Q(\theta) > Q(\theta_T)\,]\ \rightarrow\ 0 \qquad ,$$

e logo, em probabilidade, os θ_T "vão passando a estar em V", ou seja, (2.1) se verifica.

Note-se finalmente que as condições da Hipótese I são suficientes, mas não necessárias para a consistência. Maiores refinamentos, que acomodem casos como o do importante estimador de momentos simulados de Pakes (1986) e Pakes and Pollard (1989), podem ser encontrados em Newey and McFadden (1994).

A obtenção da normalidade assintótica exige mais propriedades das funções objetivo. Lembrando que $Q_T(\ \theta_T\ )$ é o valor, no ponto de ótimo, da função objetivo, para uma dada amostra de tamanho T. Suporemos que

Hipótese II (HII). A partir de um T* fixo, as funções Q_T , $T>T^*$, são de classe C^2, em uma vizinhança U de θ_0 .

---

[3] O leitor com uma formação mais analítica pode tentar construir um contra exemplo, no caso unidimensional $\Theta = R$, que mostra porque a compacidade é necessária.

Com essa hipótese, para tamanhos de amostra suficientemente grandes, pode-se expandir por Taylor a função $DQ_T=\partial Q_T/\partial\theta$ em uma vizinhança de θ_0. A consistência e a condição de primeira ordem autorizam a escrever (novamente para um T suficientemente grande):

$$DQ_T(\theta_T)' = 0 = DQ_T(\theta_0)' + D^2Q_T(\theta_T\#)\,(\theta_T - \theta_0)$$

onde $\theta_T\#$ é uma combinação convexa de θ_T e θ_0. Devido à continuidade de D^2Q_T em θ_0 e, ainda, à consistência do estimador, é possível encontrar um T** a partir do qual D^2Q_T seja inversível em uma vizinhança contida em U, o que permite escrever

$$\sqrt{T}\,(\theta_T - \theta_0) = -\,[D^2Q_T(\theta_T\#)]^{-1}\sqrt{T}\,DQ_T(\theta_0)' \qquad . \qquad (2.3)$$

Para controlar as derivadas segundas é necessário algo como

Hipótese III (HIII). As matrizes estocásticas $D^2Q_T(\theta_T\#)$ convergem em probabilidade, uniformemente em Θ, a uma matriz determinística H, simétrica e positiva definida.

Com essa hipótese, se um TCL for aplicável a $DQ_T(\theta_0)'$, o resultado estará obtido, sendo a variância assintótica da normal limite

$$H^{-1}\,\mathrm{varass}(DQ_T(\theta_0)')\,H^{-1} \qquad . \qquad (2.4)$$

No nível de generalidade dessa discussão, o resultado fica um pouco abstrato demais. De fato, hipóteses um pouco mais realistas necessitam um mínimo de informação sobre a estrutura das $Q_T(\theta)$, o que será exemplificado na seção 4.

Para terminar, há uma versão dos estimadores de extremos que guarda estreita relação com o MGM. Consiste ela no caso em que existe uma função $\varphi: \Theta \to R^q$, $q \geq p$, tal que $\varphi(\theta)$ é estimável por uma amostra de tamanho T. Chamando F_T esse estimador (que, em princípio, não depende de θ), e δ uma distância qualquer em R^q, a função objetivo é definida como

$$Q_T(\theta) = \delta\,(F_T, \varphi(\theta)) \quad . \tag{2.5}$$

Esses estimadores são chamados de distância mínima e, caso δ seja uma distância euclideana, estimadores de χ^2 mínimo. Exploraremos mais adiante as suas relações com o MGM.

3. O MGM : definição e propriedades básicas.

O MGM pode ser introduzido através duas estruturas estatísticas que, embora semelhantes, não são conceitualmente idênticas. A primeira enfatiza a existência de um modelo estatístico subjacente, ao passo que a segunda prioriza uma visão dinâmica do problema.

Sejam os vetores aleatórios Y e X, n e m-dimensionais, respectivamente, e um parâmetro $\theta \in \Theta$, sendo Θ um subconjunto do R^p. Supõe-se conhecida uma função g: $R^{n+m} \times \Theta \rightarrow R^q$, $q \geq p$, que, para cada θ fixado, dá origem ao vetor aleatório $g(Y,X;\theta)$. Existe um único $\theta_0 \in \Theta$, informalmente considerado como o "verdadeiro valor do parâmetro", para o qual

$$g(Y,X\,;\,\theta_0) \equiv 0\ .$$

O problema que o MGM se propõe a resolver é o de como a partir de uma amostra de tamanho T, Y_t, X_t, $t = 1,...,T$, da população (Y,X) se pode estimar este θ_0.

Seguindo convenções usuais em econometria, o leitor deve talvez já ter pensado em Y e X como vetores de endógenas e exógenas, resp., e na g como definindo um modelo. Ressalte-se também que, embora indexando as observações por t, elas não precisam necessariamente ser ao longo do tempo, podendo dizer respeito a um painel, uma seção reta sobre unidades (agentes) econômicas ou geográficas, ou qualquer outro tipo de amostra. No contexto do MGM, prefere-se ver a g como definindo q condições, ditas condições de momento, por razões que ficarão claras a seguir. Uma delas é que a condição de que $g(Y,X;\theta_0)$ seja identicamente nulo pode ser muito forte, sendo abrandada para

$$E\ g(Y,X\ ;\ \theta_0) = 0\ \ .$$

Ao se escrever a identidade acima uma questão suplementar advém: sob qual medida de probabilidade está sendo tomada a esperança ? Pode ocorrer que tal medida, caracterizada pela distribuição do vetor (Y,X), seja independente do valor do parâmetro, caso em que a questão não cabe. Entretanto, em geral, θ indexará também essas distribuições, sendo então necessário rescrever a igualdade como:

$$E_0\ g(Y,X\ ;\ \theta_0) = 0\ \ ,$$

onde o subscrito chama a atenção para o fato de que a esperança é calculada segundo a distribuição associada a θ_0 .

Uma outra forma de introduzir o método é considerar a existência de um processo estocástico $\{(Y_t,\ X_t)\}$ que dá origem a um outro processo (de “erros”), q-dimensional e estacionário de primeira ordem,

$$e_t = g(Y_t,X_t\ ;\ \theta) \qquad , \quad \text{tal que, para um único } \theta_0 \in \Theta\ ,$$

$$E_0\ e_t = E_0\ g(Y_t,X_t\ ;\ \theta_0) = 0 \qquad\qquad . \qquad\qquad (3.1)$$

O problema é então estimar θ_0 a partir de um número finito, $t = 1,...,T$, de observações do processo.

A ligação entre as duas estruturas é evidente. Como, na primeira, g é “identicamente nula (ou possui média zero) para o verdadeiro valor do parâmetro”, supõe-se que as observações satisfaçam aproximadamente a esta condição, que é preservada naturalmente na média. Daí a idéia de encarar, na segunda, e_t como um vetor de "erros".

Para obter o estimador MGM, constrói-se

$$g_T(\{ Y_t, X_t \}; \theta) \equiv g_T(\theta) = \sum_t g(Y_t, X_t ; \theta) / T = \sum_t g_t(\theta) / T \quad . \tag{3.2}$$

É razoável pensar que o θ que minimizará alguma norma do vetor $g_T(\theta)$ - preferencialmente até o igualará a zero - será um estimador conveniente de θ_0 . Esta é a idéia do MGM que, como se vê, generaliza a proposta de Pearson mencionada na Introdução. Tomando-se uma matriz A qxq , simétrica e semidefinida positiva, o estimador MGM é então definido como:

$$\theta_{MGM} = \arg\min_{\theta\in\Theta} g_T(\theta)' A \, g_T(\theta) \tag{3.3}$$

É comum exigir-se $A=W^{-1}$, onde W seria agora simétrica e positiva, e, também, escolher-se a ponderação de acordo com a amostra, de modo a gerar uma seqüência de matrizes estocásticas A_T como ponderadores.

A primeira propriedade do estimador definido em (3.3) é a sua consistência. Para obtê-la, supõe-se que a seqüência A_T é tal que

Hipótese 1 (H1). Plim $A_T = A$, onde A é uma matriz *determinística*, simétrica, positiva definida.

Como o limite dos ponderadores já está resolvido por H1, basta agora traduzir a Hipótese I utilizada na seção anterior para provar a consistência do estimador de extremos. Se valer

Hipótese 2 (H2). Verifica-se uma lei fraca dos grandes números *uniformemente* em Θ , para a coleção de seqüências $\{ g(Y_t, X_t ;\theta), \theta \in \Theta \}$.

poderemos escrever:

$$\text{Plim } g_T(\theta) = \lim_{T\to\infty} \sum_t E_\theta\, g\,(Y_t,X_t\,;\theta) / T = m(\theta)\,, \tag{3.4}$$

e, consequentemente,

$$\text{Plim } g_T(\theta)'\, A_T\, g_T(\theta) = m(\theta)'\, A\, m(\theta) = Q(\theta)\,. \tag{3.5}$$

Note-se que, em (3.4), o que importa é o comportamento da seqüência aleatória $\{g\,(Y_t, X_t\,;\theta)\}$ e não o do processo estocástico $\{(Y_t,X_t)'\}$. Em uma situação onde esse último não seja, por exemplo, estacionário, mas uma LfGN seja válida para a seqüência, o limite será válido. Esta, por seu turno, como sugerido no segundo membro de (3.4), não precisa necessariamente também ser estacionária, bastando, por exemplo, satisfazer à uma condição de Kolmogorov. Caso seja estacionária, como se sabe:

$$E_\theta\, g\,(Y_t,X_t\,;\theta) = E_\theta\, g\,(Y,X\,;\theta) = m(\theta)\quad . \tag{3.6}$$

O limite da função objetivo em (3.5) é, por hipótese, válido uniformemente em Θ. Entretanto, de forma análoga à seção anterior, a passagem para o limite dos pontos de mínimo necessita ainda que a função $m(\theta)$ seja contínua e que se imponha, sobre ela, a seguinte condição de identificação local:

Hipótese 3 (H3). Para o “verdadeiro valor” θ_0, $m(\theta_0) = 0$ e existe uma vizinhança V^* de θ_0, tal que $m(\theta) \neq 0$, $\forall\, \theta \in V^*$, $\theta \neq \theta_0$.

Com esta hipótese adicional, está assegurada a identificação[4] e, como A é definida positiva, o leitor concordará que a consistência está obtida, valendo a

[4] Note-se que, local se H3 valer para outros $\theta \in \Theta$, e global se somente para θ_0.

Proposição 1'. Sob as hipóteses H1, H2 e H3, e sendo $m(\theta)$ contínua sobre Θ, compacto, o problema do MGM está identificado e o estimador por ele definido é consistente.

Para obter a variância assintótica, impõe-se, além das condições para a consistência, a

Hipótese 4 (H4). As funções $g_T(\theta)$ são de classe C^1 em Θ.

Sob H4, faz sentido a condição de primeira ordem:

$$2\, [\partial g_T/\partial\theta]'\, A_T\, g_T(\theta) = 0\ . \tag{3.7}$$

Ademais, abrindo-se por Taylor $g_T(\theta)$ em uma vizinhança de θ_0 , devido à consistência, tem-se para o ponto de mínimo θ_T :

$$g_T(\theta_T) = g_T(\theta_0) + [\partial g_T/\partial\theta\#]\, (\theta_T - \theta_0) \quad , \tag{3.8}$$

onde $\theta\#$ é uma combinação convexa entre θ_T e θ_0. Levando-se em conta (3.7), pode-se escrever:

$$[\partial g_T/\partial\theta_T]'\, A_T\, g_T(\theta_0) + [\partial g_T/\partial\theta_T]'\, A_T\, [\partial g_T/\partial\theta\#]\, (\theta_T - \theta_0) = 0$$

de tal forma que:

$$\theta_T - \theta_0 = - \left([\partial g_T/\partial\theta_T]'\, A_T\, [\partial g_T/\partial\theta\#] \right)^{-1} [\partial g_T/\partial\theta_T]'\, A_T\, g_T(\theta_0) \ , \tag{3.9}$$

sendo a inversibilidade da matriz entre parênteses garantida (assintóticamente) pela consistência, como se verá a seguir. De fato, torna-se agora necessário assegurar o comportamento dos jacobianos. Para isto, de modo quase análogo ao feito para as seqüências $\{g\,(Y_t,X_t\,;\theta)\}$, suporemos que,

Hipótese 5 (H5). Uniformemente em uma vizinhança de θ_0 , vale uma LfGN para $\{\partial g(Y_t,X_t\ ;\theta)/\partial\theta\}$; podendo então ser definida, para cada θ na dita vizinhança, uma matriz determinística e de posto pleno:

$$\text{Plim } [\partial g_T/\partial\theta] = D(\theta) \quad .$$

Este fato, mais a consistência, permite dizer que:

$$\text{Plim } [\partial g_T/\partial\theta_T] = \text{Plim } [\partial g_T/\partial\theta\#] = D(\theta_0) \quad .$$

Com isto, por (3.8), a variância assintótica de $\sqrt{T}\,(\theta_T - \theta_0)$ será dada por (3.10):

$$[D(\theta_0)'AD(\theta_0)]^{-1} D(\theta_0)'A\ [\ \text{Plim } \textstyle\sum_t g(Y_t,X_t;\theta_0).(\ \sum_t g(Y_t,X_t;\theta_0))'/T\ ]\ A\ D(\theta_0)\ [D(\theta_0)'AD(\theta_0)]^{-1}$$

aparecendo então mais uma convergência, crucial, que leva a definir a matriz Φ como

$$\Phi = \text{Plim } \textstyle\sum_t (g(Y_t, X_t;\theta_0)).(\ \sum_t g(Y_t, X_t;\theta_0))'/T =$$

$$= \lim_{T\to\infty} E_0\ [\ (\ \textstyle\sum_t g(Y_t, X_t;\theta_0)\ )\ .\ (\ \sum_t g(Y_t, X_t;\theta_0)\ )'\ /\ T\ ]\ . \qquad (3.11)$$

A matriz acima merece uma inspeção cuidadosa. Suponhamos, de início, que as observações sejam independentes. Nesse caso, tendo em vista as condições de momento, obtém-se:

$$\Phi = \lim_{T\to\infty} \textstyle\sum_t E_0\ [\ g(Y_t, X_t;\theta_0)\ .\ g(Y_t, X_t;\theta_0)'\ ]\ /\ T \quad ,$$

o que, caso as esperanças cruzadas sejam estacionárias, resulta em

$$\Phi = E_0\ (\ g(Y_1, X_1;\theta_0)\ .\ g(Y_1, X_1;\theta_0)'\ )$$

$$= \text{Plim } \textstyle\sum_t g(Y_t,X_t;\theta_0).\ g(Y_t,X_t;\theta_0)'/T \quad . \qquad (3.12)$$

Ou seja, a matriz Φ é a matriz de covariâncias do processo$\{e_t\}$, ou a *matriz de covariâncias (marginais) entre as condições de momento.*

Se os índices t denotarem o tempo, e o processo (ou as observações) $\{e_t\}$ for estacionário de segunda ordem, Φ é nada mais do que a matriz de *covariâncias a longo-*

prazo do processo. Como se sabe, tal matriz é igual ao limite de T vezes a variância da média amostral, ou à matriz de densidade espectral do processo $\{e_t\}$, na freqüência zero, multiplicada por 2π[5]. Dado que o vetor $\sqrt{T}\, g_T(\theta)$ pode ser escrito como

$$(1/\sqrt{T}) \sum_t g(Y_t, X_t ; \theta_0) = (1/\sqrt{T}) \sum_t e_t \quad ,$$

Φ pode também ser vista como o limite do produto externo do vetor $(\sum_t e_t)/\sqrt{T}$ por ele mesmo.

Entretanto, a expressão (3.11) pode assumir feições outras. Supondo-se que os índices digam respeito a unidades espaciais, com uma estrutura específica de correlações entre as condições relativas a unidades adjacentes, a forma de Φ resultará de tal estrutura. A situação pode se complicar mais se os índices se referirem a observações sobre as unidades e evoluindo no tempo, combinando as duas circunstâncias discutidas.

Com essas ressalvas, pode-se enunciar a

Proposição 2: Sendo Θ compacto, sob as hipóteses H1 a H5, e existindo o limite em (3.11), a variância assintótica do estimador MGM é dada por

$$[D(\theta_0)'AD(\theta_0)]^{-1} D(\theta_0)' A \,\Phi\, A\, D(\theta_0)\, [D(\theta_0)'AD(\theta_0)]^{-1} \quad . \qquad (3.13)$$

As fórmulas (3.10) e (3.11) permitem ainda reduzir o grau de arbitrariedade na escolha do(s) ponderador(es). De fato, se as observações forem independentes ou, se ao longo do tempo, estacionárias de segunda ordem, é possível obter a expressão do ponderador otimal. Basta que os A_T sejam escolhidos de forma que

$$A_T = k \,.\, [\,\mathrm{var}\ (\sum_t g(Y_t, X_t ; \theta_0)\ /\ \sqrt{T}\,)\,]^{-1}$$

$$\cong k \,.\, [\,\sum_t (g(Y_t, X_t;\theta_0)).(\sum_t g(Y_t, X_t;\theta_0))'/T\,]^{-1} \quad , \qquad (3.14)$$

---

[5] Os leitores não familiarizados com estes conceitos devem consultar, por exemplo, a seção 10.5 de Hamilton (1994). A multiplicação, ou não, por 2π depende de como se define a densidade espectral.

ou por qualquer outra estimativa tal que $A_T \rightarrow A = k \,.\, \Phi^{-1}$, sendo k uma constante positiva qualquer. Nesse caso, não é muito difícil mostrar que a variância assintótica se reduz a

$$[\, D(\theta_0)' \, \Phi^{-1} \, D(\theta_0)]^{-1} \quad . \tag{3.15}$$

Para convencer-se de que este valor é o "menor possível", note-se primeiro que, para A qualquer:

$$[D(\theta_0)' \, A \, D(\theta_0)]^{-1} \, D(\theta_0)' \, A \, \Phi \, A \, D(\theta_0) \; [D(\theta_0)' \, A \, D(\theta_0)]^{-1} - [\, D(\theta_0)' \, \Phi^{-1} \, D(\theta_0)]^{-1} =$$

$$[D(\theta_0)' A D(\theta_0)]^{-1} D(\theta_0)' A \{\Phi - D(\theta_0) \, [D(\theta_0)' \Phi^{-1} D(\theta_0)]^{-1} \, D(\theta_0)'\} \, AD(\theta_0) \, [D(\theta_0)' AD(\theta_0)]^{-1}$$

logo, para que a diferença acima seja semidefinida positiva, o que importa é o termo central, envolvendo a matriz Φ. Como esta é positiva definida, a sua inversa também o será, possuindo uma raiz quadrada ψ ; o que permite escrever esse termo como:

$$\psi^{-1} \, [\, I_q - \psi D(\theta_0) \, [D(\theta_0)' \psi \,. \psi \, D(\theta_0)]^{-1} \, D(\theta_0)' \psi \;\,] \, \psi^{-1} = \; \psi^{-1} \, [\, I_q - P_{\psi \, D(0)} \,]^{-1} \, \psi^{--1}$$

onde

$$P_{\psi \, D(0)} = \psi D(\theta_0) \, [D(\theta_0)' \psi \,. \psi \, D(\theta_0)]^{-1} \, D(\theta_0)' \psi$$

é um operador projeção sobre o espaço gerado pelas colunas de $\psi D(\theta_0)$. Consequentemente, o novo termo entre colchetes é a projeção no espaço complementar e, como ψ^{--1} é positiva definida, o resultado desejado segue. Pode-se, então, enunciar a

Proposição 3: Sob as hipóteses da Proposição 2 e se for possível construir uma seqüência de matrizes estocásticas $\{A_T\}$ tal que $\text{Plim } A_T = k. \, \Phi^{-1}$, então, a variância assintótica do estimador MGM implementado por uma tal seqüência é mínima, sob a ordem de Loewner[6], sendo dada por (3.15).

Note-se que (3.14) não é muito precisa sobre como implementar as A_T, assunto não trivial nos casos práticos e que será tratado em outro capítulo.

Modificando-se um pouco a condição de identificação H3, e sob H4 e H5, é possível obter-se, no caso de observações iid, interessantes interpretações associadas a (3.15). Seja, inicialmente, a

Hipótese 3* (H3*). Para o "verdadeiro valor" θ_0 , $m(\theta_0) = 0$ e existe uma vizinhança V* de θ_0 , tal que $m(\theta) < 0$ (ou $m(\theta) > 0$), $\forall\ \theta \in V^*, \theta \neq \theta_0$.

Chamemos de $f(u;\theta)$ a densidade associada a uma observação $u = (Y_t, X_t)$, suposta diferenciável em θ. Para um t qualquer, pode-se escrever (todas as integrais são sobre o suporte da $f(u;\theta_0)$):

$$\int \partial\, [g(u\,;\,\theta)\,.\,f(u;\theta)]/\partial\theta_0\,.\,du =$$

$$= \int \partial g(u;\,\theta)/\partial\theta_0\,.\,f(u;\theta_0).du + \int g(u;\,\theta_0)\,.\,\partial f(u;\theta)/\partial\theta_0\,.\,1/f(u;\theta_0)\,.\,f(u;\theta_0).du$$

$$= E_0\,[\partial g(u;\,\theta)/\partial\theta_0] + E_0\,[g(u;\,\theta_0)\,.\,\partial \ln f(u;\theta)/\partial\theta_0]$$

$$= D(\theta_0) + E_0\,[g(Y_t, X_t;\,\theta_0)\,.\,s(Y_t, X_t;\,\theta_0)] \qquad (3.16)$$

onde $s(Y_t, X_t;\, \theta_0) = s_t(\theta_0) = \partial \ln f(Y_t, X_t;\, \theta)/\partial\theta_0$ é o vetor aleatório de *scores* associado à densidade. Sob condições de regularidade[7], a diferencial pode trocar com a integral, no membro à esquerda do desenvolvimento, tendo-se, por H3*:

---

[6] Lembramos que, na Teoria das Matrizes, a relação de ordem de Loewner sobre o conjunto de matrizes reais simétricas de dimensão n é definida por: $A \geq B \Leftrightarrow A\text{-}B$ é definida semipositiva. Note-se que a ordem de Loewner é uma ordem parcial.

[7] Uma versão de tais condições seria que $|\,g(u\,;\,\theta)\,.\,f(u;\theta)\,| < w(u)$, com $\int w(u)du < \infty$.

$\int \partial\, [g(u\,;\theta)\,.\,f(u;\theta)]/\partial\theta_0\,.\,du = \partial\, [\int g(u\,;\theta)\,.\,f(u;\theta).du\,]\,/\partial\theta_0$

$$= \partial\, E_\theta\, g(u\,;\theta)\,/\partial\theta_0 = 0 \quad ,$$

o que, em virtude de (3.16), fornece:

$$D(\theta_0) = -\, E_0\, [g(Y_t, X_t; \theta_0)\,.\,s(Y_t, X_t; \theta_0)] \quad . \qquad (3.17)$$

A relação acima chama-se a *Igualdade Generalizada da Informação no MGM.* Substituindo-se (3.17) em (3.15), tem-se que, sob as hipóteses em pauta, o inverso da matriz de variância assintótica será

$$E_0\, [s_t(\theta_0)'g_t(\theta_0)']\, (E_0\, [g_t(\theta_0)\,.\,g_t(\theta_0)'])^{-1}\;\; E_0\, [g_t(\theta_0)\, s_t(\theta_0)] \quad . \qquad (3.18)$$

Considerando-se o espaço de Hilbert das variáveis aleatórias de quadrado integrável definidas sobre a estrutura probabilística em questão, a diagonal de (3.18) contém os quadrados dos tamanhos das projeções de cada componente do vetor de *scores* sobre o subespaço de dimensão q gerado pelas (variáveis aleatórias das) condições de momento.

4. A normalidade assintótica do estimador MGM.

A obtenção da normalidade assintótica permite uma interessante avaliação de até onde se ganha com tratamentos gerais como o esboçado na seção 2, para os estimadores de extremos. De fato, no contexto estrito do MGM, ela pode ser obtida de (3.9) adicionando-se basicamente uma hipótese que seria:

Hipótese 6 (H6). Vale um TCL para a seqüência $\{g(Y_t\,, X_t\,;\theta_0)\}$.

Entretanto, caso se deseje aproveitar o arcabouço geral definido naquela seção[8] torna-se necessário, irônicamente, exigir mais propriedades das condições de momento. O ponto de partida passa, como visto, a ser a função objetivo Q_T. No MGM, o valor dessa função, no ponto de ótimo, para uma dada amostra de tamanho T, é escrito como

$$Q_T(\theta_T) = \min_\theta g_T(\theta)' A_T g_T(\theta) \qquad ,$$

sendo então necessário acrescentar hipóteses que assegurem HII e HIII:

Hipótese II* (HII*). A partir de um T* fixo, as funções $g_T(\theta)$, T>T*, são de classe C^2, em uma vizinhança U de θ_0.

Hipótese III* (HIII*). Vale, uniformemente em Θ, uma LfGN para as seqüências de matrizes estocásticas $\{D^2 g_t(\theta)\}$, de modo que as matrizes estocásticas $D^2Q_T(\theta_T\#)$ convergem em probabilidade, uniformemente em Θ, a uma matriz determinística H, simétrica e positiva definida.

Com essas hipóteses, para tamanhos de amostra suficientemente grandes, pode-se expandir por Taylor – de forma idêntica ao que foi feito na seção 2 – a função $DQ_T = \partial Q_T/\partial\theta$ em uma vizinhança de θ_0, sendo possível encontrar um T** a partir do qual D^2Q_T seja inversível em uma vizinhança contida em U, o que permite escrever

$$\sqrt{T}\,(\theta_T - \theta_0) = -\,[D^2Q_T(\theta_T\#)]^{-1} \sqrt{T}\, DQ_T(\theta_0)' \qquad , \qquad (4.1)$$

que nada mais é do que (2.3). Entretanto, é possível agora avançar um pouco mais. Notando-se que $DQ_T(\theta)'$ é dado pelo membro à esquerda em (3.7), e que, portanto

$$D^2Q_T(\theta) = 2\,[\partial g_T/\partial\theta]'\, A_T\, [\partial g_T/\partial\theta] + \text{“parcela com as segundas derivadas de } g_T(\theta)\text{”} \qquad ;$$

[8] Essa é a opção adotada, por exemplo, em Davidson and MacKinnon (1993), cap. 17.

para $D^2Q_T(\theta)$ calculado nos pontos $\theta_T\#$, tomando-se o limite quando $T\rightarrow\infty$, como a convergência em probabilidade dos jacobianos é, por H5, uniforme em uma vizinhança de θ_0, tem-se, pela consistência

$$\text{Plim } 2\,[\partial g_T/\partial\theta_T\#]'\,A_T\,[\partial g_T/\partial\theta_T\#] \;=\; 2\,D(\theta_0)'\,A\,D(\theta_0) \quad .$$

Ademais, a parcela envolvendo as derivadas segundas vai a zero, pois ela contém os $g_T(\theta_T\#)$ que, por (3.4), a consistência e H3, tendem à condição de ortogonalidade. Logo,

$$\text{Plim } [D^2Q_T(\theta_T\#)]^{-1} = [2\,D(\theta_0)'\,A\,D(\theta_0)]^{-1} = H^{-1} \quad . \qquad (4.2)$$

O jacobiano DQ_T' no ponto θ_0 é um vetor coluna cujas p componentes, novamente por (3.7), são somas ponderadas de T parcelas, todas elas centradas (pois são as condições de momento). Como os pesos, dados por

$$2\,[\partial g_T/\partial\theta_0]'\,A_T\,,$$

se estabilizam em $2\,D(\theta_0)A$ quando $T\rightarrow\infty$, a validade de um TCL para $\sqrt{T}\,DQ_T(\theta_0)$ depende unicamente da seqüência $\{g(Y_t\,, X_t\,;\,\theta_0)\}$, o que é assegurado por H6.

Com essas hipóteses, o membro esquerdo de (4.1) converge para uma normal de média zero e matriz de covariância dada por

$$H^{-1}\,(2D(\theta_0)'\,A)\,\Phi\,(\,2A\,D(\theta_0))\,H^{-1} \quad , \qquad (4.3)$$

onde Φ é o limite da variância de $\sqrt{T}\,g_T(\theta_0)$ definido em (3.11).

Considerando a expressão (4.2) para H, a variância assintótica pode ser reescrita como

$$[D(\theta_0)'\,A\,D(\theta_0)]^{-1}\,D(\theta_0)'\,A\,\Phi\,A\,D(\theta_0)\;\;[D(\theta_0)'\,A\,D(\theta_0)]^{-1} \qquad (4.4)$$

que, como esperado, é exatamente a que figura na Proposição 2, além de ser a tradução de (2.4) no contexto do MGM.

Tais resultados constituem a

Proposição 4. Sob as hipóteses H1 a H6 (adicionando-se, se quizer, HII* e HIII*), e sendo Θ compacto, $\sqrt{T}\,(\theta_{MGM} - \theta_0)$ converge fracamente a uma normal com média zero e variância dada por (4.4).

5. Algumas interpretações clássicas e um exemplo.

Para encerrar este capítulo, três interpretações, ligando o MGM aos métodos de estimação paramétrica convencionais, são importantes:

a) *Número de condições igual à dimensão do parâmetro – o Método dos Momentos de Pearson.* Se p=q, (3.3) se simplifica e o estimador MGM é definido pela solução do sistema

$$g_T(\theta_{MGM}) = 0 \qquad (5.1)$$

não tendo nenhuma relevância o ponderador. Isto implica que (3.8) passa a

$$0 = g_T(\theta_0) + [\partial g_T/\partial\theta\#]\,(\theta_{MGM} - \theta_0) \qquad , \qquad (5.2)$$

e como $[\partial g_T/\partial\theta\#]$ é, agora, uma matriz quadrada, a variância assintótica, sob H2 a H5, será imediatamente expressa por (3.15).

Se q>p , a condição de primeira ordem (3.7) para se obter o estimador, utilizando-se uma aproximação F_T^{-1} ao ponderador ótimo:

$$[\partial g_T/\partial\theta]'\, F_T^{-1}\, g_T(\theta) = 0$$

diz que as q condições iniciais, combinadas linearmente pela matriz $[\partial g_T/\partial\theta]' F_T^{-1}$ viram uma nova função $g_T{}^*(\theta) = [\partial g_T/\partial\theta]'\, F_T^{-1}\,(g_T(\theta))$, com exatamente p dimensões, que permite obter um estimador para o parâmetro conforme a idéia original de Pearson traduzida por (5.1). Estas p linhas do sistema $g_T{}^*(\theta) = 0$ são chamadas os *momentos ótimos.*

b) *O estimador MQO.* Seja, sob as hipóteses usuais, o problema da regressão linear, dada uma amostra de tamanho T:

$$Y_t = X_t'\theta_0 + e_t \quad , \; E\,(e_t|\, X_t) = 0 \; , \quad t = 1, 2, ..., T \; .$$

Como os regressores não são correlacionados com o erro, as condições de momento são

$$E\,(e_t.X_t) = E\,((Y_t - X_t'\theta_0)\, X_t) = 0 \; ,$$

e o estimador MGM é um caso particular de (5.1)

$$\Sigma_t\,(X_t\,(Y_t - X_t'\theta_{MGM})) = (\Sigma_t\, X_t\, Y_t) - (\Sigma_t\, X_t\, X_t')\,\theta_{MGM} = 0$$

ou, usando a notação usual,

$$\theta_{MGM} = \theta_{MQO} = (X'X)^{-1}X'Y \quad . \qquad (5.3)$$

No caso iid, com erros homocedásticos, $E(e_t)=\sigma^2$, a variância assintótica, dada por (3.15), será aproximada, como esperado, por

$$[(X'X)\;(\,E\,X'ee'X)^{-1}\,(X'X)]^{-1} = \sigma^2(X'X)^{-1} \quad .$$

Sob outras hipóteses, entretanto, a visão pelo MGM permite tratar, de forma mais simples e geral, a questão da variância assintótica. Tais situações, no contexto das variáveis instrumentais, serão vistas no Capítulo 3.

c) *Máxima verossimilhança (MV) como caso particular do MGM.* Suponha-se que o problema admita uma função de verossimilhança $L_T(\theta)$, advinda da família de densidades $f(\theta)$, e sejam as q=p condições

$$g_T(\theta) = (1/T)[\partial \log L_T/\partial\theta]' \; .$$

Pelo visto em a), independentemente do ponderador A_T , o estimador MGM para estas condições coincide com o estimador MV correspondente. Aquele terá, como variância assintótica,

$$[\;\text{Plim}\;(1/T)\;[\partial^2 \log L_T/\partial\theta^2]'\;\Phi^{-1}\;\text{Plim}\;(1/T)\;[\partial^2 \log L_T/\partial\theta^2]\;]^{-1} \quad .$$

Sob a hipótese iid e valendo as condições de regularidade da teoria de MV, tem-se

$$\Phi = E\ (\ [\partial \log f(\theta)/\partial\theta_0]'[\partial \log f(\theta)/\partial\theta_0]\ ) = -\ E[\partial^2 \log f(\theta)/\partial\theta_0^2] =$$

$$= -\ \text{Plim}\ (1/T)\ [\partial^2 \log L_T/\partial\theta^2]$$

recuperando-se o resultado que diz ser a variância assintótica do estimador de MV o inverso da matriz de informação de Fisher. Note-se que as condições da MV satisfazem H3*, valendo pois a Igualdade Generalizada da Informação provada na seção 3. Ora, nesse caso, ela não é nada mais do que as duas (conhecidas) expressões no meio das igualdades acima, uma vez que as condições de momento e o *score* coincidem.

Estes resultados são importantes para a análise de questões de otimalidade do estimador MGM, pois mostram que dadas q>p condições, o “melhor” estimador MGM, que atinge (3.15), não é necessariamente eficiente. A pergunta natural que surge é de então porque utilizar o MGM ? A resposta é tripla. Primeiro, como sabido, em que pese os desenvolvimentos da pseudo-verossimilhança, o estimador de MV pode rapidamente perder a sua eficiência quando a hipótese distribucional é violada. Segundo, a solução do programa dado por (3.3) pode ser computacionalmente mais simples do que a maximização da verossimilhança escolhida[9]. A obtenção analítica da própria verossimilhança pode ser muito difícil, senão impossível. Finalmente, a hipótese iid pode ser violada, destruindo as propriedades ótimas da MV.

O MGM é, portanto, uma alternativa mais flexível que deve ser usada sempre que houver incerteza sobre a hipótese distribucional ou esta se revelar pouco tratável. Nas aplicações econométricas, ele é um convite a desenvolver as condições de momento a partir da teoria subjacente ao problema. Ajunte-se ainda que, devido à sua flexibilidade, ele se adapta mais facilmente a um contexto de estimação semiparamétrica - técnica que vem ganhando relevância atualmente.

---

[9] Um exemplo sofisticado disto pode ser encontrado em Potscher and Prucha (1986).

Finalmente, para concluir o capítulo, seja o

Exemplo. Trata-se de um caso tradicional, ligado à formulação de Pearson, que é muito bem desenvolvido em Greene (2000)[10]. Considerando-se uma amostra iid de uma v.a. X com densidade gama dada por

$$f(x) = \lambda^{\alpha} e^{-\lambda x} x^{\alpha-1} \;/\; \Gamma(\alpha) \qquad ; \quad x>0 \;,\; \alpha, \lambda > 0 \quad ,$$

tem-se, por propriedades elementares, as seguintes condições de momento:

$$EX^k = \lambda^{-k}\, \Gamma(\alpha+k) \;/\; \Gamma(\alpha) \qquad , \quad k=1,2, \ldots$$

$$E \ln X = d \ln\Gamma(\alpha) / d\alpha \;-\; \ln\lambda$$

$$E\, 1/X = \lambda \;/\; \alpha\text{-}1 \qquad .$$

Tomando-se os dois primeiros momentos não-centrados e as duas outras igualdades, obtem-se quatro condições para os dois parâmetros. Para usá-las simultaneamente, é necessário um estimador do ponderador ótimo. Utiliza-se então uma estimativa inicial através da primeira e da terceira condição (ElnX) ; estimativa esta que coincide com o resultado da estimação por MV[11].

Para obter a variância assintótica do estimador MGM, denotando por

$g_T(\alpha,\lambda) =$

$$= [\; \lambda(\Sigma X_i/T)-\alpha \quad \lambda^2(\Sigma X_i^2/T)-\alpha(\alpha+1) \quad (\alpha-1)(\Sigma X_i^{-1}/T)-\lambda \quad (\Sigma \ln X_i/T)-\Psi(\alpha)-\ln\lambda \;]\,'$$

o vetor com as quatro condições de momento (sendo $\Psi(\alpha) = d\ln\Gamma(\alpha) \,/\, d\alpha$), é necessário calcular $\partial g_T(\alpha,\lambda)/\partial(\alpha,\lambda)$, o que dará uma matriz 4x2 onde aparecerá a derivada segunda da função gama.

Suponha-se agora que, utilizando-se as quatro condições da maneira esboçada acima, encontra-se estimativas para os desvios-padrão dos estimadores razoavelmente inferiores às encontradas para o caso da MV[12]. Isto é estranho, pois

[10] Capítulos 4 e 11.
[11] O leitor deve verificar esta afirmação.
[12] Tal é o caso no exemplo mencionado de Greene (2000).

sabe-se que os estimadores de MV são eficientes. Provavelmente, o teste sobre a validade das quatro condições - a ser discutido em um próximo capítulo - rejeitará esta hipótese nula, indicando, no caso, que *talvez a hipótese distribucional adotada não seja válida.* É importante notar que, se o pesquisador recorresse diretamente à MV, êle não teria a possibilidade de fazer este teste de especificação. Isto só foi possível pelo uso das condições *adicionais* de momento – isto é, uso de mais informação - que o MGM permite.

Referências

Burguete, J. F. 1980. Asymptotic theory of instrumental variables in nonlinear regression. Ph.D. Dissertation, North Carolina State University, Raleigh, North Carolina.

Davidson, R. and J. G. MacKinnon. 1993. *Estimation and Inference in Econometrics*. Oxford University Press, New York.

Gallant, A. R. 1977. Three-stage least-squares estimation for a system of simultaneous, nonlinear, implicit equations. *Journal of Econometrics* 5, 71 - 88.

Gallant, A. R. and D. W. Jorgenson. 1979. Statistical inference for a system of simultaneous, nonlinear implicit equations in the context of instrumental variables estimation. *Journal of Econometrics* 11, 275 - 302.

Greene, W. H. 2000. *Econometric Analysis.* 4th edition. Prentice Hall Inc., New Jersey.

Hamilton, J. 1994. *Time Series Analysis*. Princeton University Press, Princeton.

Hansen, L. P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029 - 54.

Manski, C. F. 1988. *Analog Estimation Methods in Econometrics.* Chapman and Hall, London.

Newey, W. K. and D. Mc Fadden. 1994. Large sample estimation and hypothesis testing, in R. F. Engle and D. L. McFadden, eds., *Handbook of Econometrics*, Vol. 4, North-Holland, Amsterdam.

Pakes, A. 1986. Patents as options: some estimates of the value of holding European patent stocks. *Econometrica* 54, 755-85.

Pakes, A. and D. Pollard. 1989. Simulation and the asymptotics of optimization estimators. *Econometrica* 57, 1027 - 57.

Potscher, B. M. and I. R. Prucha. 1986. A class of partially adaptive one-step M-estimators for the nonlinear regression model for dependent observations. *J. of Econometrics* 32, 219-51.

471. CUSTO DE CICLO ECONÔMICO NO BRASIL EM UM MODELO COM RESTRIÇÃO A CRÉDITO - Bárbara Vasconcelos Boavista da Cunha; Pedro Cavalcanti Ferreira – Janeiro de 2003 – 21 págs.

472. THE COSTS OF EDUCATION, LONGEVITY AND THE POVERTY OF NATIONS - Pedro Cavalcanti Ferreira; Samuel de Abreu Pessoa – Janeiro de 2003 – 31 págs.

473. A GENERALIZATION OF JUDD'S METHOD OF OUT-STEADY-STATE COMPARISONS IN PERFECT FORESIGHT MODELS - Paulo Barelli; Samuel de Abreu Pessoa – Fevereiro de 2003 – 7 págs.

474. AS LEIS DA FALÊNCIA: UMA ABORDAGEM ECONÔMICA - Aloísio Pessoa de Araújo – Fevereiro de 2003 – 25 págs.

475. THE LONG-RUN ECONOMIC IMPACT OF AIDS - Pedro Cavalcanti G. Ferreira; Samuel de Abreu Pessoa – Fevereiro de 2003 – 30 págs.

476. A MONETARY MECHANISM FOR SHARING CAPITAL: DIAMOND AND DYBVIG MEET KIYOTAKI AND WRIGHT – Ricardo de O. Cavalcanti – Fevereiro de 2003 – 16 págs.

477. INADA CONDITIONS IMPLY THAT PRODUCTION FUNCTION MUST BE ASYMPTOTICALLY COBB-DOUGLAS - Paulo Barelli; Samuel de Abreu Pessoa – Março de 2003 – 4 págs.

478. TEMPORAL AGGREGATION AND BANDWIDTH SELECTION IN ESTIMATING LONG MEMORY - Leonardo R. Souza - Março de 2003 – 19 págs.

479. A NOTE ON COLE AND STOCKMAN - Paulo Barelli; Samuel de Abreu Pessoa – Abril de 2003 – 8 págs.

480. A HIPÓTESE DAS EXPECTATIVAS NA ESTRUTURA A TERMO DE JUROS NO BRASIL: UMA APLICAÇÃO DE MODELOS DE VALOR PRESENTE - Alexandre Maia Correia Lima; João Victor Issler – Maio de 2003 – 30 págs.

481. ON THE WELFARE COSTS OF BUSINESS CYCLES IN THE 20TH CENTURY - João Victor Issler; Afonso Arinos de Mello Franco; Osmani Teixeira de Carvalho Guillén – Maio de 2003 – 29 págs.

482. RETORNOS ANORMAIS E ESTRATÉGIAS CONTRÁRIAS - Marco Antonio Bonomo; Ivana Dall'Agnol – Junho de 2003 – 27 págs.

483. EVOLUÇÃO DA PRODUTIVIDADE TOTAL DOS FATORES NA ECONOMIA BRASILEIRA: UMA ANÁLISE COMPARATIVA - Victor Gomes; Samuel de Abreu Pessoa;Fernando A . Veloso – Junho de 2003 – 45 págs.

484. MIGRAÇÃO, SELEÇÃO E DIFERENÇAS REGIONAIS DE RENDA NO BRASIL - Enestor da Rosa dos Santos Junior; Naércio Menezes Filho; Pedro Cavalcanti Ferreira – Junho de 2003 – 23 págs.

485. THE RISK PREMIUM ON BRAZILIAN GOVERNMENT DEBT, 1996-2002 - André Soares Loureiro; Fernando de Holanda Barbosa - Junho de 2003 – 16 págs.

486. FORECASTING ELECTRICITY DEMAND USING GENERALIZED LONG MEMORY - Lacir Jorge Soares; Leonardo Rocha Souza – Junho de 2003 – 22 págs.

487. USING IRREGULARLY SPACED RETURNS TO ESTIMATE MULTI-FACTOR MODELS: APPLICATION TO BRAZILIAN EQUITY DATA - Álvaro Veiga; Leonardo Rocha Souza – Junho de 2003 – 26 págs.

488. BOUNDS FOR THE PROBABILITY DISTRIBUTION FUNCTION OF THE LINEAR ACD PROCESS – Marcelo Fernandes – Julho de 2003 – 10 págs.

489. CONVEX COMBINATIONS OF LONG MEMORY ESTIMATES FROM DIFFERENT SAMPLING RATES - Leonardo R. Souza; Jeremy Smith; Reinaldo C. Souza – Julho de 2003 – 20 págs.

490. IDADE, INCAPACIDADE E A INFLAÇÃO DO NÚMERO DE PESSOAS COM DEFICIÊNCIA - Marcelo Neri ; Wagner Soares – Julho de 2003 – 54 págs.

491. FORECASTING ELECTRICITY LOAD DEMAND: ANALYSIS OF THE 2001 RATIONING PERIOD IN BRAZIL - Leonardo Rocha Souza; Lacir Jorge Soares – Julho de 2003 – 27 págs.

492. THE MISSING LINK: USING THE NBER RECESSION INDICATOR TO CONSTRUCT COINCIDENT AND LEADING INDICES OF ECONOMIC ACTIVITY - JoãoVictor Issler; Farshid Vahid – Agosto de 2003 – 26 págs.

493. REAL EXCHANGE RATE MISALIGNMENTS - Maria Cristina T. Terra; Frederico Estrella Carneiro Valladares – Agosto de 2003 – 26 págs.

494. ELASTICITY OF SUBSTITUTION BETWEEN CAPITAL AND LABOR: A PANEL DATA APPROACH - Samuel de Abreu Pessoa ; Silvia Matos Pessoa; Rafael Rob – Agosto de 2003 – 30 págs.

495. A EXPERIÊNCIA DE CRESCIMENTO DAS ECONOMIAS DE MERCADO NOS ÚLTIMOS 40 ANOS – Samuel de Abreu Pessoa – Agosto de 2003 – 22 págs.

496. NORMALITY UNDER UNCERTAINTY – Carlos Eugênio E. da Costa – Setembro de 2003 – 08 págs.

497. RISK SHARING AND THE HOUSEHOLD COLLECTIVE MODEL - Carlos Eugênio E. da Costa – Setembro de 2003 – 15 págs.

498. REDISTRIBUTION WITH UNOBSERVED 'EX-ANTE' CHOICES - Carlos Eugênio E. da Costa – Setembro de 2003 – 30 págs.

499. OPTIMAL TAXATION WITH GRADUAL LEARNING OF TYPES - Carlos Eugênio E. da Costa – Setembro de 2003 – 26 págs.

500. AVALIANDO PESQUISADORES E DEPARTAMENTOS DE ECONOMIA NO BRASIL A PARTIR DE CITAÇÕES INTERNACIONAIS - João Victor Issler; Rachel Couto Ferreira – Setembro de 2003 – 29 págs.

501. A FAMILY OF AUTOREGRESSIVE CONDITIONAL DURATION MODELS - Marcelo Fernandes; Joachim Grammig – Setembro de 2003 – 37 págs.

502. NONPARAMETRIC SPECIFICATION TESTS FOR CONDITIONAL DURATION MODELS - Marcelo Fernandes; Joachim Grammig – Setembro de 2003 – 42 págs.

503. A NOTE ON CHAMBERS'S "LONG MEMORY AND AGGREGATION IN MACROECONOMIC TIME SERIES" – Leonardo Rocha Souza – Setembro de 2003 – 11págs.

504. ON CHOICE OF TECHNIQUE IN THE ROBINSON-SOLOW-SRINIVASAN MODEL - M. Ali Khan – Setembro de 2003 – 34 págs.

505. ENDOGENOUS TIME-DEPENDENT RULES AND THE COSTS OF DISINFLATION WITH IMPERFECT CREDIBILITY - Marco Bonomo; Carlos Viana de Carvalho – Outubro de 2003 – 27 págs.

506. CAPITAIS INTERNACIONAIS: COMPLEMENTARES OU SUBSTITUTOS? - Carlos Hamilton V. Araújo; Renato G. Flôres Jr. – Outubro de 2003 – 24 págs.

507. TESTING PRODUCTION FUNCTIONS USED IN EMPIRICAL GROWTH STUDIES - Pedro Cavalcanti Ferreira; João Victor Issler; Samuel de Abreu Pessoa – Outubro de 2003 – 8 págs.

508. SHOULD EDUCATIONAL POLICIES BE REGRESSIVE ? Daniel Gottlieb; Humberto Moreira – Outubro de 2003 – 25 págs.

509. TRADE AND CO-OPERATION IN THE EU-MERCOSUL FREE TRADE AGREEMENT - Renato G. Flôres Jr. – Outubro de 2003 – 33 págs.

510. OUTPUT CONVERGENCE IN MERCOSUR: MULTIVARIATE TIME SERIES EVIDENCE - Mariam Camarero; Renato G. Flôres Jr; Cecílio Tamarit – Outubro de 2003 – 36 págs.

511. ENDOGENOUS COLLATERAL - Aloísio Araújo; José Fajardo Barbachan; Mario R. Páscoa – Novembro de 2003 – 37 págs.

512. NON-MONOTONE INSURANCE CONTRACTS AND THEIR EMPIRICAL CONSEQUENCES - Aloísio Araujo; Humberto Moreira – Novembro de 2003 – 31 págs.

513. EQUILIBRIA IN SECURITY MARKETS WITH A CONTINUUM OF AGENTS - A. Araujo; V. F. Martins da Rocha; P. K. Monteiro – Novembro de 2003 – 17 pág's.

514. SPECULATIVE ATTACKS ON DEBTS AND OPTIMUM CURRENCY AREA: A WELFARE ANALYSIS - Aloisio Araujo; Márcia Leon – Novembro de 2003 – 50 págs.

515. O MÉTODO GENERALIZADO DOS MOMENTOS(MGM): CONCEITOS BÁSICOS - Renato G. Flôres Jr – Novembro de 2003 – 27 págs.

516. VARIÁVEIS INTRUMENTAIS E O MGM: USO DE MOMENTOS CONDICIONAIS - Renato G. Flôres Jr – Novembro de 2003 – 27 págs.

517. O VALOR DA MOEDA E A TEORIA DOS PREÇOS DOS ATIVOS - Fernando de Holanda Barbosa – Dezembro de 2003 – 17 págs.

518. EMPRESÁRIOS NANICOS, GARANTIAS E ACESSO À CRÉDITO - Marcelo Côrtes Néri; Fabiano da Silva Giovanini - Dezembro de 2003 – 23 págs.

519. DESENHO DE UM SISTEMA DE METAS SOCIAIS - Marcelo Côrtes Néri; Marcelo Xerez - Dezembro de 2003 – 24 págs.

520. A NEW INCIDENCE ANALYSIS OF BRAZILIAN SOCIAL POLICIES USING MULTIPLE DATA SOURCES - Marcelo Côrtes Néri - Dezembro de 2003 – 55 págs.